DOCUMENT RESUME

ED 458 273 TM 033 436

AUTHOR Brown, Pamela J.; Augustine, Andy
TITLE Screen Reading Software as an Assessment Accommodation: Implications for Instruction and Student Performance.
PUB DATE 2001-04-00
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Accommodations (Disabilities); *Computer Software; Educational Testing; *High School Students; High Schools; Sciences; Social Studies; *Test Format; Testing Problems
IDENTIFIERS *Screen Readers

ABSTRACT

Whether assessment items administered using screen reading software measure students learning better than assessment items in a paper-and-pencil format was studied. Using a computer to present a test orally controls for standardization of administration and allows each student to complete the assessment at his/her own pace. In this study, 96 students completed a science assessment and 110 completed a social studies assessment. One version was administered in the traditional paper-and-pencil format while the other version was administered through a computer using screen reading software. To compare student performance on the two versions of the assessment, a repeated-measures design using the general linear model was used. The results of the repeated-measures analysis of covariance reveal that for both the social studies and science assessments, the students' reading score had a significant effect. However, format (screen reading versus paper/pencil) did not have a significant impact on the scores on this assessment when controlling for a student's reading ability. While this study revealed no significant differences between the performance of students completing the pencil-and-paper format version versus the screen reading format when controlling for reading performance, using screen reading software as an accommodation in science for students with poor reading skills might still be effective. It is likely that the lack of significant results are compounded by the lack of appropriate instruction for students with poor reading skills. That is, if reading is the primary instructional method for students to learn concepts in the content areas of science and social studies, then students who performed poorly on these assessments, performed poorly because of lack of knowledge about science or social studies rather than inability to comprehend the test questions. (Contains 5 tables and 21 references.) (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 458 273

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as
received from the person or organization
originating it.
Minor changes have been made to
improve reproduction quality.
Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

P.J. Brown

1

SCREEN READING SOFTWARE AS AN ASSESSMENT ACCOMMODATION: IMPLICATIONS FOR INSTRUCTION AND STUDENT PERFORMANCE

Pamela J. Brown, Ph.D., Associate Policy Scientist
Andy Augustine, M.A., Educator-in-Residence

Delaware Education Research & Development Center
University of Delaware
Newark, DE 19716

TM033436

BEST COPY AVAILABLE

Paper Presented at the American Education Research Association Annual Meeting, Seattle, WA
April 2001

EXECUTIVE SUMMARY

The purpose of this research study was to determine if assessment items administered using screen reading software measure student learning better than assessment items in a paper and pencil format. Using a computer to present a test orally controls for standardization of administration and allows each student to complete the assessment at his/her own pace. Few published studies have used a computer to present a test orally (Burk, 1998).

In this study, 96 students completed a science assessment and 110 students completed a social studies assessment. One version was administered in the traditional paper and pencil format while the other version was administered via a computer utilizing screen reading software. The purpose of this study was to determine if the format of the assessment (screen reading vs. paper/pencil) differentially affected student performance. In order to compare student performance on the two versions of the assessment, a repeated-measures design using the general linear model (GLM) was used.

The results of the repeated-measures ANCOVA revealed that for both the social studies and the science assessment, the students' reading score had a significant effect. However, format (screen reading versus paper/pencil) did not have a significant impact on the scores on this assessment when controlling for a student's reading ability.

While this study revealed no significant differences between the performance of students completing the pencil and paper format version versus the screen reading format when controlling for reading performance, using screen reading software as an accommodation in science for students with poor reading skills might still be effective. It is likely that the lack of significant results are compounded by the lack of appropriate instruction for students with poor reading skills. That is, if reading is the primary instructional method for students to learn

concepts in the content areas of science and social studies, then students who performed poorly on these assessments, performed poorly because of lack of knowledge about science or social studies rather than inability to comprehend the test questions.

INTRODUCTION

The purpose of this research study was to determine if assessment items administered using screen reading software measure student learning better than assessment items in a paper and pencil format. This study is part of a larger study entitled the Inclusive Comprehensive Assessment System (ICAS) Project. The goal of the ICAS project is to evaluate various assessment methods or accommodations that maximize access to large-scale assessments by eliminating barriers in testing situations that are not relevant to the construct being measured. This study is specifically designed to evaluate the usefulness of screen reading software for assessments for students with reading difficulties as well as those without reading difficulties.

Several research studies on the K-12 student population have focused on the use of computer-based testing (CBT) which generally involves using a computer to administer a paper and pencil test (Burk, 1998; Curtis & Kropp, 1961; Hasselbring & Crossland, 1982; Horton & Lovitt, 1994; Keene & Davey, 1987; Miller, 1990; Swain, 1997; Varnhagen & Gerber, 1984; and Watkins & Kush, 1988). Other studies on the K-12 student population have focused on presenting the tests using audio cassettes, video cassettes, or human readers (Bennett, Rock, & Kaplan, 1987; Epsin & Sindelar, 1988; Harker & Feldt, 1993, Helwig, Tedesco, Heath, Tindal, & Almond, 1998; Koretz, 1997; Tachibana, 1986; Tindal, Almond, Heath, & Tedesco, 1998; Tindal, Glasgow, Helwig, Hollebeck, & Heath, 1998; Tindal, Heath, Hollenbeck, Almond, & Harniss, 1998; Trimbal, 1998; Westin, 1999). The studies that explore the use of audio or video cassettes in a classroom permit a standard administration of the assessment. On the other hand, these devices generally are administered to an entire class of students and thus do not allow individual students to work at their own pace. Using a human reader also does not allow individual students to work at their own pace. In addition, using a human reader also presents

other problems such as a lack of standardization of the assessment administration. Using a computer to present a test orally controls for standardization of administration and allows each student to complete the assessment at his/her own pace. Few published studies, however, have used a computer to present the test orally (Burk, 1998).

METHODOLOGY

Creation of the Assessments

For this study, four assessments were created and administered -- two in the area of social studies and two in the area of science. The assessments were comprised of publicly released NAEP (National Assessment of Educational Progress) items that were selected by several experienced Delaware and Pennsylvania high school social studies and science teachers. Items on both versions of the assessment were matched for content area, process skill, and difficulty level assessed. In addition, the items were arranged in order of difficulty from the easiest to the most difficult.

Participants Selected

For this study, eighteen school districts in Delaware and three school districts in Pennsylvania were contacted to participate. Eleven high schools across eight school districts throughout Delaware and two school districts in Pennsylvania agreed to participate. Consent forms were distributed to all high school seniors (n = 2,593) as well as to their parents in each of these schools. Less than one-fourth (13.6%) of the parents and students returned the consent forms after two mailings. Most parents (74.2%) who returned the consent forms gave their consent, but some of these students were unable to participate due to absenteeism or withdrawal from school. The sample included students who had reading difficulties (as measured by a standardized reading test) as well as students that did not have reading difficulties. Table 1

contains information about the reading level of the participants. For Delaware students their 10th grade Delaware Student Testing Program (DSTP) reading score was used to determine their reading level.

Table 1
Reading Level of Students (as measured by national standardized tests) Who Completed the Assessment by Content Area

Content	Range of Reading Percentile	Mean Reading Percentile	Standard deviation	Total Sample Size
Science	5-99	57.23	26.88	96
Social Studies	1-99	55.08	27.08	110

Research Design

To ensure that there were no order effects, the design was counter-balanced. That is, half of the students began with Version A and finished with Version B while the other half began with Version B and finished with Version A. Also, half of the students began with the paper/pencil format while the other began with the screen reading format. Table 2 presents the research design used.

Administration of the Assessments

Ninety-six students completed the science assessment and 110 students completed the social studies assessment. Each version consisted of a variety of grade-appropriate multiple choice and open-ended items (see Appendix A). One version was administered in the traditional paper and pencil format while the other version was administered via a computer utilizing screen reading software. Authorware 5.0 was the software package used for the administration of the screen reading portion of this study. All students completed both versions of the assessment so

as to serve as their own control for this study. This controls for the impact of extraneous variables such as race, gender, age, and SES on the results of this study.

Table 2
Number of Students Selected to Participate in Research Study

Content Area	Format Completed First	
	Paper/Pencil	Screen Reading
Social Studies	50	50
Version A in paper/pencil format AND Version B in screen reading format	25	25
Version A in screen reading format AND Version B in paper/pencil format	25	25
Science	50	50
Version A in paper/pencil format AND Version B in screen reading format	25	25
Version A in screen reading format AND Version B in paper/pencil format	25	25
Total	100	100

Screen reading software permitted the student to listen via a headset to the test items as they were displayed on the computer screen. Each student could choose to listen to any assessment item multiple times. Students selected an answer for the multiple-choice items by using the mouse to click on option A, B, C, or D. For the open-ended items, students typed their answer into a text box on the screen.

Each correct response to a multiple choice item received one point while the open-ended item was scored using a 3-point or 4-point rubric. A total score was calculated by summing the scores received for each item on the assessment. The total score was also converted to a percentage correct score. Table 3 provides a summary of the type of items on each assessment administered.

The purpose of this study was to determine if the format of the assessment (screen reading vs. paper/pencil) differentially affected student performance. In order to compare student performance on the two versions of the assessment, a repeated-measures design using the general linear model (GLM) was used. The within- subjects factor was the students' scores on the assessments. There was no between subjects factor for this study. The percentile rank on the reading portion of a national standardized test served as the covariate. Furthermore, a regression analysis was conducted to determine if a student's reading score was useful in predicting a student's science or social studies assessment score.

Table 3
Description of Mathematics and Science Assessments Administered

	Version	Number of Items	Type of Items	Total Score Possible
Social Studies	A	5	Open-Ended	15
		13	Multiple Choice	13
	B	5	Open-Ended	16
		12	Multiple Choice	12
Science	A	2	Open-Ended	6
		31	Multiple Choice	31
	B	2	Open-Ended	6
		30	Multiple Choice	30

Scoring Process for the Open-Ended Items

Each open-ended item was scored by a rater using the rubric that accompanied the NAEP assessment item. The raters for the items had strong backgrounds in the appropriate content area. Since the rubrics were straightforward (see Figures 1 & 2), only one rater was used to score each item. However, to control for bias, the same rater scored all assessments for a given item.

Figure 1.
Example of Scoring Rubric for a Science Item

3 = Complete - student response describes two ways in which heart disease can be prevented, such as those below. 2 = Partial - student response describes one way in which heart disease can be prevented. 1 = Unsatisfactory/Incorrect - student response shows no understanding of how heart disease can be prevented.
Credited responses include: getting more exercise, regular exercise; reducing stress/relaxing, eating less saturated fat/avoid greasy food

Figure 2.
Example of Scoring Rubric for a Social Studies Item

3 = Appropriate - These answers explain the link between a factor and suburbanization, citing specifics or elaborating on the explanation. 2 = Partial - These answers suggest a linkage between a factor and suburbanization, but it is vague and lack specifics. 1 = Inappropriate - These answers do not address the linkage between a factor and the growth of suburbs.
Credited responses could include: - automobiles and highways enabled people to move further away from places where they work and shopped, encouraging the growth of communities (suburbs) at some distance from the workplace, from which people can commute. - tax deductions enabled more people to buy homes, which lead to the rapid growth of suburban areas (sprawl).

Reliability Analysis

In the tables below is a summary of the reliability statistics for the two versions of the social studies and science assessments. Reliability statistics are given for each assessment as a whole. Since there are fewer items on the social studies assessment than the science assessment, one would expect lower reliability statistics on the social studies assessments.

Table 3
Reliability Statistics (Coefficient Alpha)

	Version A	Version B
Social Studies Assessment	.79	.71
Science Assessment	.87	.83

FINDINGS

The results of the repeated-measures ANCOVA revealed that for both the social studies and the science assessment, the students' reading score (covariate) had a significant effect. However, format (screen reading versus paper/pencil) did not have a significant impact on the scores on these assessments when controlling for a student's reading ability. The results of these tests are shown in Tables 4 - 5.

Table 4
ANCOVA for a Repeated-Measures Design for the Social Studies Assessment

Source	df	F
Between Subjects		
Intercept	1	234.19**
Reading	1	38.46**
error	83	(291.09)
Within Subjects		
Test Score	1	.67
Test Score*Reading	1	.04
error	83	(128.51)

Note. Values enclosed in parentheses represent mean square errors.
** $p < .01$

Table 5
ANCOVA for a Repeated-Measures Design for the Science Assessment

Source	df	F
	Between Subjects	
Intercept	1	170.89**
Reading	1	23.11**
error	84	(432.12)
	Within Subjects	
Test Score	1	.57
Test Score*Reading	1	.43
error	84	(103.30)

Note. Values enclosed in parentheses represent mean square errors.
** $p < .01$

While there were no significant differences between formats, there were significant differences between the scores of good readers and the scores of struggling readers (see Tables 6 & 7). In addition, there were significant differences between poor and average readers on the social studies assessment in both formats. This supports the hypothesis that on average good readers perform better than poor readers in the science and social studies.

Table 6
Descriptive Statistics on the Science Assessment for Good, Average, and Struggling Readers

Format	Good Readers n=37	Average Readers n=26	Poor Readers n=23
Paper/Pencil	72.12 (18.33)	66.15 (16.34)	53.51 (15.88)
Screen Reading	71.77 (16.41)	63.16 (19.05)	57.08 (14.24)

Note: For this study, good readers were defined as those students scoring above the 67th percentile and struggling readers were defined as those students scoring below the 34th percentile. Average readers were defined as those falling between the 34th and 67th percentiles.

Table 7

Descriptive Statistics on the Social Studies Assessment for Good, Average, and Struggling Readers

Format	Good Readers n=29	Average Readers n=37	Poor Readers n=19
Paper/Pencil	71.80 (13.06)	65.06 (12.80)	50.75 (12.63)
Screen Reading	66.13 (20.44)	62.74 (13.46)	46.43 (15.38)

Note: For this study, good readers were defined as those students scoring above the 67th percentile and struggling readers were defined as those students scoring below the 34th percentile. Average readers were defined as those falling between the 34th and 67th percentiles.

To illuminate these findings, a regression analysis was also conducted. The regression analysis revealed that for the social studies assessment as well as the science assessment, the students reading score was a significant predictor of their performance. Those students who had high reading scores tended to score well on these assessments regardless of the format. In the case of the social studies assessment, this regression model predicts almost 27% of the variance of the scores. With the science assessment, this model predicts about 19% of the variance of the scores. The results of these analyses are presented in Tables 8 and 9.

Table 8

Summary of Regression Analysis for Variables Predicting Total Score on Social Studies Assessment

Variable	B	SE B	β
Reading Percentile	.073	.015	.470**
Version (A or B)	-3.02	3.74	-.08
Format (Paper/Pencil or Screen Reading)	-1.16	.82	-.14

Note. $R^2 = .266$, ** $p < .01$

Table 9
Summary of Regression Analysis for Variables Predicting Total Score on Science Assessment

Variable	B	SE B	ß
Reading Percentile	.260	.060	.437**
Version (A or B)	-.89	3.24	-.03
Format (Paper/Pencil or Screen Reading)	-2.69	3.20	-.08

Note. $R^2 = .190$, ** $p < .01$

SUMMARY

This study revealed no significant differences between the performance of students completing the pencil and paper format version versus the screen reading format when controlling for reading performance. However, it is likely that the limited numbers of significant results are compounded by the lack of appropriate instruction for students with poor reading skills. That is, if reading is the primary instructional method for students to learn concepts in the content areas of science and social studies, then students who performed poorly on these assessments may have performed poorly because of lack of knowledge about science or social studies rather than their inability to comprehend the test questions. To tease out this factor (primary method of instruction), one would need to secure a sample of students who have been instructed using methods that do not require the students to learn primarily by reading.

References

Bennett, R. E., Rock, D. A., & Kaplan, B. A. (1987). SAT differential items performance for nine handicapped groups. *Journal of Education Measurement,* 24(1), 44-55.

Burk, M. (1998, October). *Computerized test accommodations: A new approach for inclusion and success for students with disabilities.* Paper presented at Office of Special Education Program Cross Project Meeting "Technology and the Education of Children with Disabilities: Steppingstones to the 21st Century.

Curtis, H. A., & Kropp, R. P. (1961). A comparison of scores obtained by administering a test normally and visually. *Journal of Experimental Education,* 29, 249-260.

Epsin, C. A., & Sindelar, P. T. (1988). Auditory feedback and writing: Learning disabled and nondisabled students. *Exceptional Children,* 55, 45-51.

Harker, J. K., & Feldt, L. S. (1993). A comparison of achievement test performance of nondisabled students under silent reading and reading plus listening modes of administrations. *Applied Measurement,* 6, 307-320.

Hasselbring, T. S., & Crossland, C. L. (1982). Application of microcomputer technology to spelling assessment of learning disabled students. *Learning Disability Quarterly,* 5, 80-82.

Helwig, R., Tedesco, M., Health, B., Tindal, G., & Almond, P. (1998). *The relationship between reading ability and performance on a video accommodated math problem-solving test.* Manuscript submitted for publication, University of Oregon.

Horton, S. V., & Lovitt, T. C. (1994). A comparison of two methods of administering group reading inventories to diverse learners. *Remedial and Special Education,* 15, 378-390.

Keene, S., & Davey, B. (1987). Effects of computer-presented text on LD adolescents' reading behaviors. *Learning Disability Quarterly,* 10, 283-290.

Koretz, D. (1997). *The assessment of students with disabilities in Kentucky* (CSE Technical Report No. 431). Los Angeles, CA: Center for Research on Evaluation, Standards and Student Testing.

Miller, S. (1998). *The relationship between language simplification of math word problems and performance for students with disabilities.* Unpublished master's project, University of Oregon, Eugene, OR.

Swain, C. R. (1997). A comparison of computer-administered test and a paper and pencil test using normally achieving and mathematically disabled young children (Doctoral dissertation, University of North Texas, 1997). *Dissertation Abstracts International,* 58, 0158.

Tachibana, K. K. (1986). Standardized testing modifications for learning disabled college students in Florida (modality) (Doctoral dissertation, University of Miami, 1986). *Dissertation Abstracts International,* 47, 0125.

Tindal, G., Almond, P., Heath, B., & Tedesco, M. (1998). *Single subject research using audio cassette read aloud in math.* Manuscript submitted for publication, University of Oregon.

Tindal, G., & Fuchs, L. (1999). *A summary of research on test changes: An empirical basis for defining accommodations.* Unpublished manuscript with Mid-South Regional Resource Center, University of Kentucky.

Tindal, G., Glasgow, A., Helwig, B., Hollebeck, K., & Heath, B. (1998). *Accommodations in large scale tests for students with disabilities: An investigation of reading math tests using video technology.* Unpublished manuscript with Council of Chief State School Officers, Washington, DC.

Tindal, G., Heath, B., Hollenbeck, K., Almond, P., & Harniss, M. (1998). Accommodating students with disabilites on large-scale tests: An emperical study of student response and test administration demands. *Exceptional Children,* 64(4), 439-450.

Trimbal, S. (1998). *Performance trends and use of accommodations on a statewide assessment* (Maryland/Kentucky State Assessment Series Rep. No. 3). Minneapolis, MN: National Center on Educational Outcomes.

Varnhagen, S., & Gerber, M. M. (1984). Use of microcomputers for spelling assessment: Reasons to be cautious. *Learning Disability Quarterly,* 7, 266-270.

Watkins, M. W., & Kush, J. C. (1988). Assessment of academic skills of learning disabled students with classroom microcomputers. *School Psychology Review,* 17, 81-88.

Westin, T. (April, 1999). *The validity of oral presentation in testing.* Montreal, Canada: American Educational Research Association.